A-I

19K41A0449

ECE-A

Assignment -7

Q) Draw a decision tree diagram to predic no. of hours to play based on weather conditions like outlook, humidity, windy consider data set shown below.

Outlook	Temperature	Humidity	windy	hours to play.
Rainy	Hot	High	false	25
Raining	Hot	High	True	30
overcast	hot	high	false	46
sunny	mild	high	false	45
Sunny	cool	Normal	True	52
sunny	cool	Normal	True	23
over cast	cool	normal	false	43
Rainy	cool	Normal	flese	38
sunny	mild	normal	tne	46
Rainy	mild	normal	True	48
over cast	mild	High	True	52
over cost	Hot	normal	flese	44
sunny	mild	high	True.	30

Termination Criterial CVC = 10% or min no. of samples = 4

Calculation of mean, standard deviation, co-efficient of variation

$$\text{mean} = \frac{\Sigma x}{n} = \frac{557}{14} = 39.78$$

$$SD = \sqrt{\frac{\Sigma(x-\text{mean})^2}{n}} = 9.67$$

$$CV = \frac{SD}{\text{mean}} \times 100 = \frac{9.67}{39.78} \times 100 = 24.30$$

Here, data set is split into different attributes. The SD of each branch is calculated

$$SD(\text{attr}) = \Sigma w(\text{branch}) \cdot SD(\text{branch})$$ & then SDR (standard ~~reduction~~ deviation reduction) is calculated.

$$SDR = SD - SD(\text{attr})$$

$$SD = 9.67$$

out look	mean	SD	CV	n	w(v)
Rainy	35.2	8.7	24.7	5	5/14
over cast	46.25	4.03	8.72	4	4/14
sunny	39.2	12.2	31.0	5	5/14

$\therefore SD(\text{out look}) = 5/14 \times 8.7 + \frac{4}{14} \times 4.03 + \frac{5}{14} \times 12.2 = 8.59$

$SDR(\text{out look}) = SD - SD(\text{out look})$

$= 9.67 - 8.59$

$= 1.08$

19Ka1A0449 (5)

Temperature

Temp	mean	SD	CV	n	ω(V)
hot	36.25	10.34	30.6	4	4/14
cool	39	12.14	31.1	4	4/14
mild	36.6	8.38	19.65	6	6/14

$\therefore SD(\text{temperature}) = \frac{4}{14} \times 10.34 + \frac{4}{14} \times 12.14 + \frac{6}{14} \times 8.38 = 10.01$

$SDR(\text{Temp}) = SD - SD(\text{temp}) = 9.67 - 10.01 = 0.34$

Humidity

Humidity	mean	SD	CV	n	ω(HU)
High	37.51	10.11	26.92	7	7/14
Normal	32	9.4	22.4	7	7/14

$\therefore SD(\text{humidity}) = 7/14 \times 10.11 + \frac{7}{14} \times 9.4 = 9.77$

$SDR(\text{humidity}) = SD - SD(\text{humidity})$

$= 9.67 - 9.77 = 0.1$

windy:

windy	mean	SD	CV	n	ω(V)
True	37.6	11.6	30.8	6	6/14
False	31.3	8.01	20.3	8	8/14

$\therefore SD(\text{windy}) = \frac{6}{14} \times 11.6 + \frac{8}{14} \times 8.01 = 9.77$

$SDR(\text{windy}) = SD - SD(\text{windy}) = 9.67 - 9.77 = -0.1$

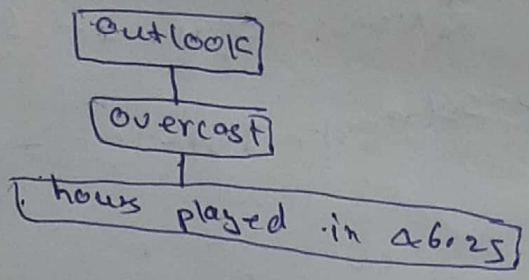
The value that has highest SDR is considered as root node (i.e, decision node)

considering termination criteria

CV is 10% or CV is (n ≤)

outlook:-

overcast has cv of 8% which is less than threshold value therefore, we need not go to further splitting



we need to split sunny & rainy columns

sunny:-

outlook	Temp	Humidity	windy	hours played
sunny	mild	high	false	45
sunny	cool	normal	false	52
sunny	cool	normal	true	23
sunny	mild	normal	false	46
sunny	mild	high	true	30

∴ mean = 39.2, SD = 12.2, CV = 31.0

Temperature

Temp	mean	SD	CV	n	w(v)
mild	40.3	8.96	22.73	3	3/5
cold	35.5	20.50	54.66	2	2/5

$SD(Temp) = \frac{3}{5} \times 8.96 + \frac{2}{5} \times 20.50 = 13.576$

$SDR(temp) = SD - SD(temp) = 12.2 - 13.576 = -1.37$

Humidity

humidity	mean	sd	cv	n	co(v)
high	37.5	10.6	28.26	2	2/5
normal	40.3	15.3	37.96	3	3/5

$SD(humidity) = \frac{2}{5} \times 10.6 + \frac{3}{5} \times 15.3 = 13.42$

$SDR(humanity) = SD - SD(humanity)$

$= 12.2 - 13.42$

$= -1.22$

Windy

windy	mean	SD	CV	n	co(v)
false	47.66	3.78	7.94	3	3/5
True	26.5	4.94	18.65	2	2/5

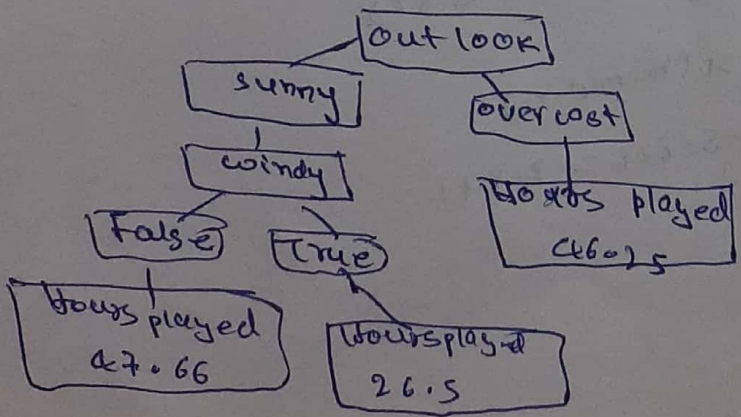
$SD(windy) = \frac{3}{5} \times 3.78 + \frac{2}{5} \times 4.94 = 4.23$

$SDR(windy) = SD - SD(windy) = 12.2 - 4.23 = 7.97$

In outlook, among temp, humidity & windy SDR value is high for windy SDR = 7.97

Then check for CV value

Both true & false satisfy the CV value



Rainy:

outlook	temp	humidity	windy	Hours played.
Rainy	Hot	high	False	25
Rainy	hot	high	True	30
Rainy	mild	high	false	35
Rainy	cool	normal	false	38
Rainy	mild.	normal	True	48

mean = 35.2, SD = 8.7, CV = 24.7

Temperature

Temp

	mean	SD	CV	n	co(v)
Hot	27.5	3.53	12.83	2	25
mild	41.5	9.19	22.144	2	215
Cool	38	0	0	1	115

$SD(temp) = \frac{2}{5} \times 3.53 + \frac{2}{5} \times 9.19 + \frac{1}{5} \times 0 = 5.088$

$SDR(temp) = SD - SD(temp) = 8.7 - 5.088 = 3.612$

Humidity

Humidity

	mean	SD	CV	n	co(v)
high	30	5	16.66	3	315
normal	43	7.07	16.44	2	215

$SD(humidity) = \frac{3}{5} \times 5 + \frac{2}{5} \times 7.07 = 5.828$

$SDR(humidity) = SD - SD(humidity)$

$= 8.7 - 5.828 = 2.872$

windy:-

windy	mean	SD	CV	n	(W)
false	32.66	6.80	20.85	3	3/5
True	39	12.72	32.5	2	2/5

$SD(windy) = \frac{3}{5} \times 6.80 + \frac{2}{5} \times 12.72 = 9.168$

$SDR(windy) = SD - SD(windy) = 8.7 - 9.168 = -0.468$

Among temp, humidity & windy the SDR value is high for temp (i.e 3.612). Then check for CV value at hot, mild, cold. Satisfy the CV value

Decision tree diagram to predict number of hours to play based on weather conditions